Remote Sensing of Environment 150 (2014) 132–139

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA

Itai Kloog [a,*], Francesco Nordio [b], Brent A. Coull [c], Joel Schwartz [b]

[a] Department of Geography and Environmental Development, Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, Israel
[b] Department of Environmental Health — Exposure, Epidemiology and Risk Program, Harvard School of Public Health, Landmark Center 401 Park Dr West, Boston, MA 02215, USA
[c] Department of Biostatistics, Harvard School of Public Health, Boston, MA, 02215, USA

ARTICLE INFO

Article history:
Received 9 January 2014
Received in revised form 24 April 2014
Accepted 26 April 2014
Available online 21 May 2014

Keywords:
MODIS
Surface temperature
Air temperature
Exposure error
Epidemiology

ABSTRACT

Air temperature (*Ta*) stations have limited spatial coverage, particularly in rural areas. Since temperature can vary greatly both spatially and temporally, *Ta* stations are often inadequate for studies on the health effects of extreme temperature and climate change. Satellites can provide us with daily physical surface temperature (*Ts*) measurements, enabling us to estimate daily *Ta*. In this study, we aimed to extend our previous work on predicting *Ta* from *Ts* in Massachusetts by predicting 24 h *Ta* means on a 1 km grid across the Northeast and Mid-Atlantic states, extending both the temporal and spatial coverage, improving upon the methodology and validating our model in other geographical regions across the Northeastern part of the USA. We used mixed model regressions to first calibrate *Ts* and *Ta* measurements, regressing *Ta* measurements against day-specific random intercepts, and fixed and random *Ts* slopes. Then to capture the ability of neighboring cells to fill in the cells with missing *Ts* values, we regress the *Ta* predicted from the first mixed effects model against the mean of the *Ta* measurements on that day, separately for each grid cell. Out-of-sample tenfold cross-validation was used to quantify the accuracy of our predictions. Our model performance was excellent for both days with available *Ts* and days without *Ts* observations (mean out-of-sample $R^2 = 0.95$ and $R^2 = 0.94$ respectively). We demonstrate how *Ts* can be used reliably to predict daily *Ta* at high resolution in large geographical areas even in non-retrieval days while reducing exposure measurement error.

1. Introduction

Spatially distributed near-surface air temperature (*Ta*) is a commonly used and important variable in various scientific fields including climatology (Smith et al., 1988; Watson & Albritton, 2001), study of vector-borne disease (Garske, Ferguson, & Ghani, 2013; Goetz, Prince, & Small, 2000) environmental and ecological studies (Zanobetti & Schwartz, 2008), hydrological studies (Wang, Koike, Yang, & Yeh, 2009) and many others. One particular field of importance is epidemiology and public health. Multiple studies in recent years have shown that *Ta* plays an important role in epidemiological studies. This growing body of evidence clearly show that *Ta* has a strong association with adverse health outcomes, especially for cardiovascular diseases (Halonen, Zanobetti, Sparrow, Vokonas, & Schwartz, 2010; Medina-Ramón, Zanobetti, Cavanagh, & Schwartz, 2006; Ren et al., 2011). Since temperature can vary greatly both spatially and temporally, *Ta* stations with limited spatial coverage are often inadequate for such epidemiological studies looking at extreme temperature (both warm and cold) and climate change.

Changes in climate will lead to warmer temperatures and more extreme weather events, which are associated with increased morbidity and mortality in sensitive populations (Ostro, Roth, Green, & Basu, 2009), both in heat waves and less extreme events (Basu, Feng, & Ostro, 2008; Ostro et al., 2009; Zanobetti & Schwartz, 2008). Most of the epidemiological studies on health effects of temperature have been conducted using large geographical areas as units of the analysis (state, nationwide), thus potentially biasing the health effects risks estimates due to exposure measurement error (i.e., assigning inaccurate exposure to each study participant) (Zeger et al., 2000).

Ta is commonly measured at a reference height of 2 m above the ground in most weather stations (NCDC, 2010), which are dependent on the regional infrastructure for weather data collection. These measurements that are collected as point samples are not optimal for capturing the spatial variability of the climatic conditions within each region (Vancutsem, Ceccato, Dinku, & Connor, 2010). The spatial information available on air temperature is thus often very limited, especially where the monitoring density is low (such as rural areas), since *Ta* is affected by properties that vary greatly in both space and time (Prihodko & Goward, 1997). These "out of coverage" rural areas lack detailed unbiased *Ta* readings and people who live in more densely populated areas are unlikely to be representative of those who do not. *Ta* sites cannot measure high spatial resolution exposure for epidemiology studies, which introduces exposure error, and likely biases the health effect estimates downward (Zeger et al., 2000). In addition, the urban heat

* Corresponding author. Tel.: +972 8 6428394.
E-mail address: ekloog@hsph.harvard.edu (I. Kloog).

http://dx.doi.org/10.1016/j.rse.2014.04.024
0034-4257/

island effect results in temperature variations over large spatial scales in urban areas depending on the relative amounts of hard surface versus green space.

This lack of high resolution continuous spatio-temporal measurements, has resulted in studies looking into methods for addressing this lack of data. Some studies have looked at interpolation techniques such as global interpolators, kriging and inverse density weighting (IDW) among others (Florio, Lele, Chang, Sterner, & Glass, 2004; Vicente-Serrano, Saz-Sánchez, & Cuadrat, 2003; Yang, Wang, & August, 2004). Vicente-Serrano et al. (2003) presents a detailed comparative analysis predicting temperature and precipitation from various statistical methods including multiple interpolation methods (such as kriging, co-kriging, and global interpolators). The study is conducted in an area with significant geographic differences and spatial climatic diversity (the middle Ebro Valley in the northeast of Spain). The accuracy of their comprehensive temperature analysis varied greatly ranging from $R^2 = 0.39$ for the co-kriging analysis and up to $R^2 = 0.75$ for the regression-based method. They conclude by stating that though there were small differences in estimation accuracy between the final regression model and the maps obtained from local interpolations, it is preferable to use the empirical regression model since it provides a more accurate reflection of thermal diversity at a local scale. In addition, regression-based methods are not as conditioned by the location of the *Ta* stations as local and geostatistical methods are, which give better predictions in areas with low spatial density of available data. Florio et al. (2004) also looked at different spatially dependent (kriging) models and multiple linear regressions to model *Ta*. Their results show that the kriging models predicted temperature better than the multiple regression models (0.9^C error). They do state however, that incorporating satellite data in their kriging models would result in better model performance under optimal day conditions. These station-based interpolation techniques suffer from arbitrary locations of weather stations (Lennon & Turner, 1995) and often lack continuous daily data accessibility.

Satellite surface temperature data (*Ts*) is defined as the skin temperature of the ground, which is the actual reading from surface level which the satellite "sees" when it looks through the atmosphere to the ground. *Ts* data can provide actual physical measurements across the entire study area which can improve daily *Ta* predictions.

In recent years researchers have started using satellite-based methods due to the ability to get spatial estimates of *Ts* at high temporal (daily) and spatial resolution (1 km) (Vancutsem et al., 2010; Zhang et al., 2013; Zhu, Lű, & Jia, 2013). Whereas *Ta* measurements provide point data at a relatively coarse resolution, satellite based *Ts* data can be used for a spatially continuous and high resolution view of the land *Ts*. The ability to obtain both high spatial and temporal resolutions *Ts* data is a significant advantage of satellite observations over traditional climatic methods.

Several studies correlated daytime *Ts* satellite observations with *Ta* at ground meteorological stations across non-urban areas. Most of the studies focus on predicting minimum *Ta*. Dousset (Dousset, 1989), has shown that *Ts* correlations with *Ta* are significantly better during the night since microscale advection is reduced. During the day hours when there is direct solar illumination, additional factors lead to more complex interaction between *Ta* and *Ts* including sky view factor, thermal properties of the underlined surfaces, satellite–sun geometry, surface geometry, and weather conditions. Both Fu et al. (2011) and Vancutsem et al. (2010) explored estimating *Ta* using MODIS *Ts* data with moderate success. For example in Fu's study a moderate predictive accuracy was presented ($R^2 > 0.55$, $p < 0.01$). Recently, Zhu et al. (2013) explored estimating daily maximum and minimum *Ta* using MODIS *Ts*. In order to improve the accuracy of the estimation of daily maximum *Ta*, they used the temperature–vegetation index (TVX) method and applied it to improve the daily maximum *Ta* accuracy. They report a RMSE value of 3.79 °C and model fits (R^2) of 0.83. Comparisons in urban areas have been few.

All previous studies either lack complete spatial and temporal predictions, present relatively low predictive power or do not predict daily 24 hour mean *Ta* predictions. We recently published a study (Kloog, Chudnovsky, Koutrakis, & Schwartz, 2012) on the temporal and spatial assessment of minimum *Ta* using MODIS *Ts* measurements in Massachusetts, USA. We generated daily *Ta* predictions for 2003 even when *Ts* data was unavailable. Our model performance was excellent for both days with available *Ts* and days without *Ts* observations (mean out-of-sample $R^2 = 0.95$ and $R^2 = 0.94$ respectively). A key difference from the previous approaches was the use of a mixed model allowing the regression coefficient between *Ts* and *Ta* to vary by day, resulting in considerably higher R^2.

In this paper we extend our previous work (Kloog et al., 2012) in Massachusetts by looking at predicting 24 h *Ta* means based on both night and day *Ts* data, extending both the temporal and spatial coverage of the model, improving upon the methodology and validating our model across the Northeastern part of the USA. Specifically, we developed new methodologies to better predict *Ta* over a large areas taking into account the considerable differences in such large geographical regions (covering the entire Northeast and Mid-Atlantic areas of the USA) while predicting mean 24 h *Ta* as opposed to minimum air temperature in our previous model. We also examine performance for day and night temperatures, and incorporate monitored *Ta* within 60 km buffers around each grid cell to fill in for days with missing *Ts*. We incorporate land use regression, and meteorological variables to predict daily 24 h mean *Ta* for grid cells even when satellite *Ts* measures are not available.

2. Methods

2.1. Study domain

Our study region included the entire Northeastern part of the USA (Fig. 1), and includes many urban areas (notably New York, Boston, Washington, DC and Philadelphia) and rural towns, large forested regions, water bodies, mountains and the Atlantic sea shoreline.

2.2. Surface temperature and emissivity

We obtained daily *Ts* data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite (Terra), with a spatial resolution of 1×1 km. We used the MOD11_A1 product (Terra land surface temperature & emissivity) used in many previous studies (Kloog et al., 2012; Wan, 2008) since its product is publicly available, provides daily *Ts* at a high spatial resolution and covers large areas across the USA. Emissivity was obtained since it's a parameter that can influence *Ts* measurements by causing a reduction of surface-emitted radiance. In addition, the anisotropy of reflectivity and emissivity may reduce or increase the total radiance from the surface (Weng, Lu, & Schubring, 2004). Emissivity is available in the MODIS MOD11_A1 product. The radiometric temperature was corrected for atmospheric transmission to account for the kinetic temperature of the object using the following formula (Jensen, 2009):

$$T_{kin} = T_{rad}/\text{Emmisivity}(\lambda)^{1/4}$$

where T_{kin} is related to the true kinetic temperature (and further denoted here as *Ts*) and T_{rad} is the radiant temperature of an object recorded by a remote sensor. More details about MODIS *Ts* data can be found in previous papers (Kloog et al., 2012; Wan, 2008).

2.3. Meteorological data

Daily data for *Ta* across the Northeast U.S. for 2000–2011 were obtained from two sources: The National Climatic Data Center (NCDC) and Weather underground Inc. (WU). The spatial coverage of NCDC stations was lacking (325 stations across all study area) and thus to

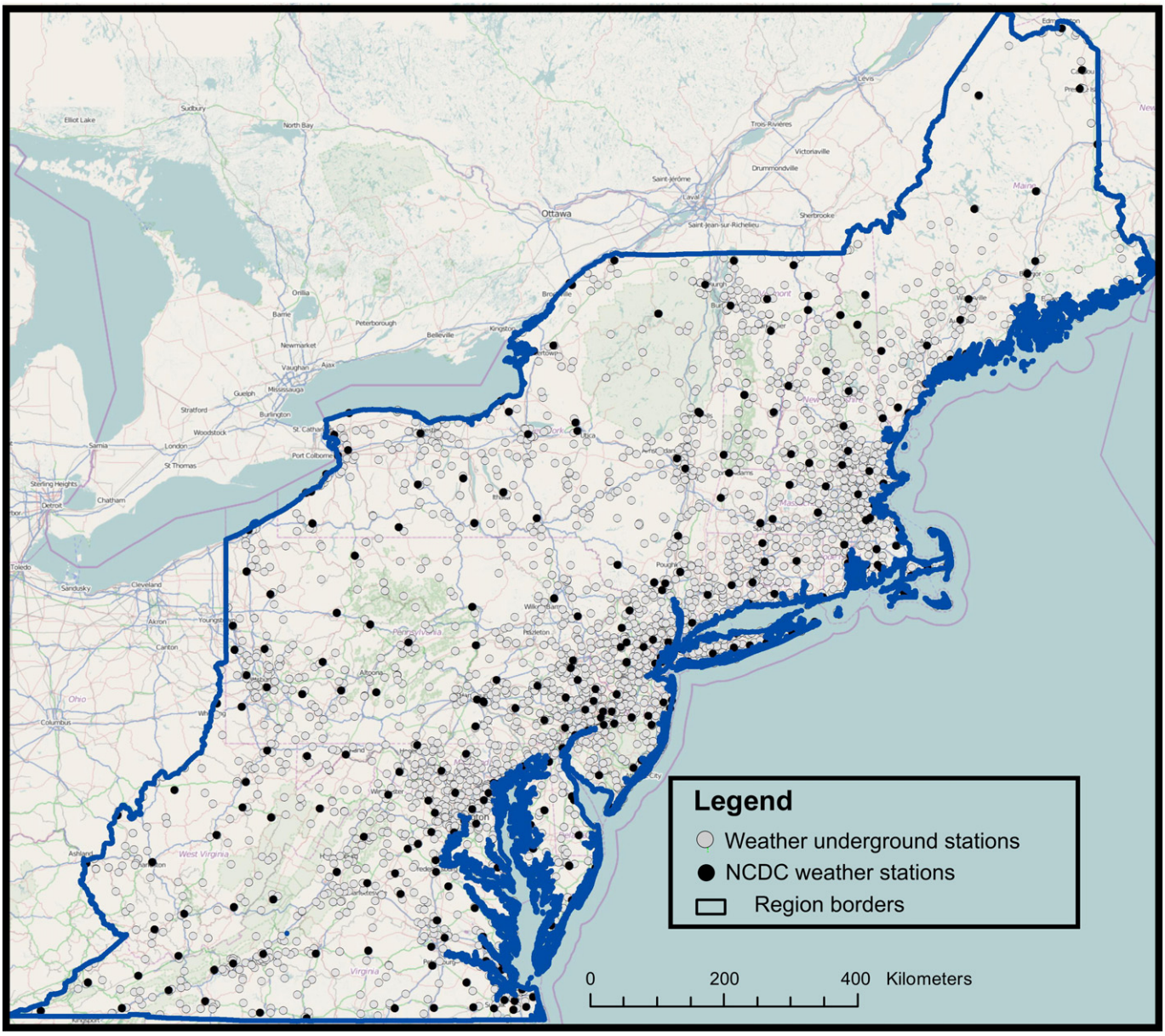

Fig. 1. Map of the study area showing all available weather underground and NCDC Air temperature monitor stations across Northeastern USA for 2000–2011.

improve the spatial coverage of *Ta* monitors we added the WU *Ta* stations to our analysis (which added additional 2535 stations). WU is a commercial provider of weather information services that incorporates a network of personal weather stations. These sources have been used in multiple studies in the past few years (Kloog et al., 2012; Von Klot, Melly, Coull, Dutton, & Schwartz, 2008). There were between 513 and 2860 daily monitors (monitors increase yearly) operating across the study area during the study period (see Fig. 1). To check differences in measurements that might result in biases, we validate the WU data comparing it to the NCDC stations by matching each of the WU stations to the closest NCDC station within 1 km, and computed the correlations. The average correlation (R^2) was 0.958, suggesting that the use of WB and WU stations is unlikely to cause bias.

2.4. Spatial predictors of air temperature

To improve the predictive ability of the final constructed model, the following spatial predictors were used in our models: normalized difference vegetation index (NDVI), percent urban and elevation.

NDVI: We used the publicly available monthly MODIS NDVI product (MOD13A3) at 1 km spatial resolution. The monthly resolution was chosen since NDVI values do not change considerably each month.
Elevation: elevation data were added through a satellite based DEM (digital elevation model) from the National Elevation Dataset (NED) (Maune, 2007). NED data is publicly available from the U.S. Geological Survey (USGS) and contains elevation data across the United States at a spatial resolution of 1 arc sec. Arc seconds were aggregated up so that each 1 × 1 km elevation cell represents the mean of that cell. There are sharp elevation contrasts across such a large study area and thus we used elevation as a spatial predictor (generally higher elevations are associated with lower air temperatures).
Percent urban: We used the national land cover data (NLCD) from 2001 (Homer, Huang, Yang, Wylie, & Coan, 2004), available as raster

files with a 30 m spatial resolution. We used values classified as built areas (NLCD values for developed low intensity, developed medium intensity and developed high intensity) to calculate the percent of urban areas in each 1 × 1 km grid across the study area.

2.5. Statistical methods

Our *Ta* predictions are generated by a series of statistical models. We start by calibrating daily *Ts* in each grid cell for which both *Ta* and *Ts* values are available (model 1). Each *Ta* stations is assigned the closest *Ts* observation on that specific day (within a distance of 1 km). On each day we estimate a separate slope in the relationship between *Ta* and *Ts* to capture the day to day temporal variability of the relationship. Random slopes are used rather than separate slopes for each day because random slopes are shrinkage estimates that reduce what otherwise could be excessive noise in day specific slopes. We performed the calibrations with both day and night MODIS *Ts* separately. Since results were almost identical (see Appendix A) we choose to use the night *Ts* as the main predictor since the quality of data is generally better. During nighttime, the earth surface behaves almost as an isothermal and homogeneous surface. In contrast, during daytime there is a significant directional anisotropy effect which is due to differences in sun illumination vs. satellite viewing geometry and different shading effects within pixels, giving rise to temperature differences as much as 20 °C (Wan & Dozier, 1996). In model 1 we fit a mixed model regression with day-specific random intercepts and *Ts* slopes as follows:

$$Ta_{ij} = \left(\alpha + u_j\right) + \left(\beta_1 + v_j\right)Ts_{ij} + \beta_2 \text{Percent urban}_i + \beta_3 \text{Elevation}_i + \beta_4 \text{NDVI}_{ij} + \varepsilon_{ij}$$
$$\left(u_j v_j\right) \sim [(o\ o), \Sigma]$$

where: Ta_{ij} is the measured *Ta* at a spatial site *i* on a day *j*; α and u_j are the fixed and random intercepts, respectively, Ts_{ij} is the *Ts* value in the grid cell corresponding to site *i* on a day *j*; β_1 and v_j are the fixed and random slopes, respectively. Percent urban is the percent of urban area in the gird cell, NDVI_{ij} is the monthly NDVI value in the grid cell corresponding to site *i* for the month in which day *j* falls. Elevation_i is the mean elevation in the grid cell corresponding to site *i*. Finally, Σ is an unstructured variance–covariance matrix for the random effects and ε_{ij} is the error term at site *i* on a day *j*.

We used ten-fold out of sample cross validation (CV) to validate our predictions at monitor locations at each step. We randomly divide our data into 90 and 10% splits ten times. We predict for the 10% data sets using the model fitted from the remaining 90% of the data. We then report these computed R^2 values. To test our results for bias we regress the measured *Ta* values against the predicted values in each site on each day. We estimated the model prediction precision by taking the square root of the mean squared prediction errors (RMSPE) as follows:

$$\text{RMSPE} = \left(N^{-1}\sum_{i=1}^{N}(Pi - Oi)^2\right)^{0.5}$$

where: *N* is the number of observations, *Oi* is the *i*th observed value and *Pi* is the *i*th predicted value.

In addition we estimated prediction errors separately for the temporal and spatial components of the model. Temporal R^2 was calculated by regressing Delta *Ta* against Delta predicted:

$$\textit{Delta Ta} \sim \textit{Delta predicted}$$

where: Delta *Ta* is the difference between the actual *Ta* in place *i* at time *j* and the annual mean *Ta* at that location, and Delta predicted is defined similarly for the predicted values generated from the model. Spatial R^2 was calculated by regressing the annual overall mean *Ta* against the mean predicted *Ta* at place *i* (each monitoring location).

We then make use of the model 1 fit to predict *Ta* in grid cells without *Ta* measurements but with available daily *Ts* satellite measurements. *Ts* values are not available in all grid/day combinations (due to retrieval errors, snow, clouds etc.) and thus this step often fails to provide predictions for many grid cell-day combinations.

To estimate *Ta* when no *Ts* data are available we fit a second model (model 2) which uses the association of predicted grid cells *Ta* values (based on satellite *Ts*) with surrounding *Ta* monitors and the association with values in neighboring grid cells. Since daily *Ta* for the whole of the Northeast varies considerably between different geographical regions, we built 60 km buffers around each *Ta* monitor. We then use the daily mean *Ta* in each buffer as a predictor of temperature in all cells that fell within the buffer. This parameter denoted mTa_{jr} is the mean *Ta* across the specific buffer *r* on a day *j*. In contrast to the first model, the second model includes a cell specific fixed slope and intercept:

$$PredTa_{ij} = \alpha_i + \beta_{i*} mTa_{jr} + \varepsilon_{ij}$$

where $PredTa_{ij}$ is the predicted *Ta* at a spatial site i on a day *j* from the first prediction model; mTa_{jr} is the mean *Ta* across the specific buffer *r* on a day *j*; α are the fixed intercepts and β_1 are the fixed slopes that varies by spatial site *i*. Statistical Analysis was performed in R (version 2.15) and SAS© (version 9.3). These coefficients were then used to fill in the predicted Ta_{ij} on days when the satellite data was missing and no monitoring data was available in these grid cells.

Finally we ran several sensitivity analyses for a sample year (2011) to see how the final model predictions perform in smaller geographic regions (such as single states and cities) and in urban areas vs. rural areas. We also ran all years to check the models performance at different time periods (summer vs. winter).

3. Results

Fig. 2 shows how applying our daily calibration method approach significantly improves the models fit. The left scatter plot shows the relationship between monitored *Ta* and satellite *Ts* before the daily calibration ($R^2 = 0.88$) in 2003. The right scatter plot shows the relationship between monitored *Ta* and the predictions from our calibration stage in 2003 (cross validated $R^2 = 0.95$).

The first stage calibration models for the Night *Ts* are presented in Table 1 (the Day *Ts* results can be found in Appendix A). The model revealed very high out-of-sample fits with a mean out of sample R^2 of 0.947 ($P < 0.001$, yearly range 0.933–0.958), and as expected a highly significant association between *Ta* and the main explanatory variable-*Ts* (Table 1). The spatial and temporal CV results also presented very good fits (Table 1). For the spatial variation in temperature the mean out-of-sample R^2 across years was 0.832 and for the temporal variability the mean out-of-sample R^2 was 0.956 ($P < 0.001$). We found no bias in our cross validation results (overall mean slope of observed vs. predicted = 1.00).

The model also yielded very small prediction errors compared to similar studies (Prihodko & Goward, 1997; Vancutsem et al., 2010; Zhu et al., 2013): our model resulted in a RMSPE of 2.157 °C and spatial RMSPE of 1.215 °C indicating strong model performance.

The second model (model 2) also performed extremely well (Table 2) with a mean R^2 across years of 0.940 ($P < 0.001$, yearly range 0.902–0.962) which is very high considering that these were days with neither ground *Ta* data nor satellite *Ts* data in the grid cells.

The sensitivity analysis results also show very good overall model performance across all analyses and are presented in Table 3. Results from validating the model per state all showed excellent model performance ranging from $R^2 = 0.892$–0.966 and RMSPE of 1.426 to 2.908. Validating our model in the greater Boston area and the NYC area resulted in a R^2 of 0.960, and 0.913 respectively and RMSPE of 1.700–2.664. We also examined the performance in rural vs. urban areas. The model preformed slightly better in urban areas as

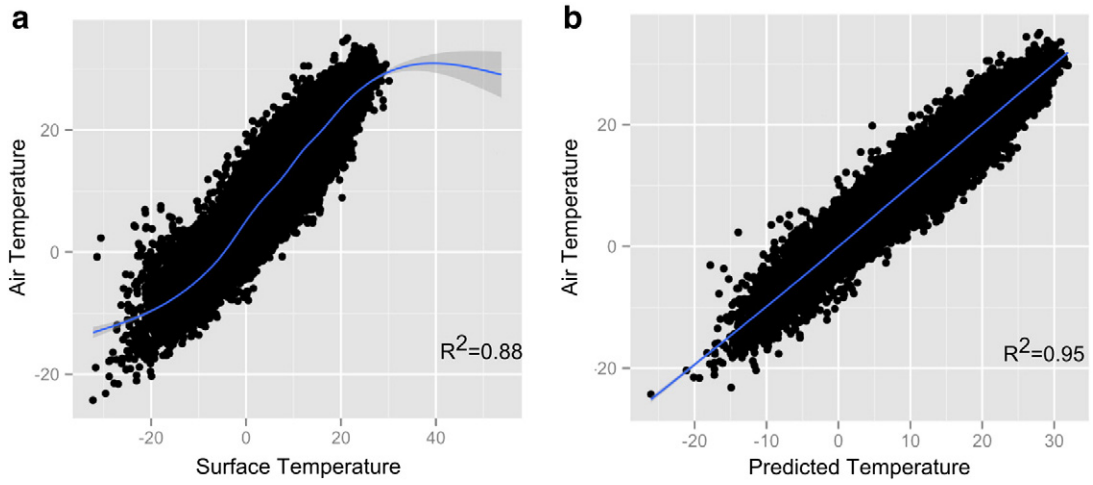

Fig. 2. A scatter plot of the air temperature–surface temperature relationship before (a) and after (b) the daily 2003 calibrations.

Table 1
Prediction accuracy: Ten-fold cross validated (CV) results for model 1 predictions (calibration stage using MODIS night surface temperature for 2000–2011).

Yearly dataset	CV R^2	Intercept [a]	Slope [a]	CV R^2 (Spatial component)	CV R^2 (Temporal component)	RMSPE[b]	RMSPE[b] (Spatial)
2000	0.933	−0.019 ± 0.022	1.001 ± 0.001	0.814	0.945	2.109	1.246
2001	0.943	0.000 ± 0.015	1.000 ± 0.001	0.806	0.958	2.102	1.183
2002	0.953	0.021 ± 0.018	0.999 ± 0.001	0.823	0.964	2.124	1.195
2003	0.958	0.022 ± 0.017	0.999 ± 0.001	0.884	0.964	2.082	1.107
2004	0.951	0.001 ± 0.014	1.000 ± 0.001	0.829	0.960	2.113	1.161
2005	0.946	0.021 ± 0.016	0.999 ± 0.001	0.800	0.954	2.219	1.225
2006	0.945	0.025 ± 0.013	0.999 ± 0.001	0.757	0.955	2.103	1.167
2007	0.945	0.020 ± 0.013	0.999 ± 0.001	0.923	0.952	2.360	1.290
2008	0.947	−0.006 ± 0.011	1.000 ± 0.001	0.828	0.954	2.141	1.238
2009	0.943	0.005 ± 0.009	1.000 ± 0.001	0.850	0.953	2.224	1.209
2010	0.949	0.013 ± 0.009	1.000 ± 0.001	0.870	0.957	2.224	1.253
2011	0.955	−0.005 ± 0.009	1.000 ± 0.001	0.806	0.958	2.083	1.303
Mean 2000–2011	0.947	0.008 ± 0.014	1.000 ± 0.001	0.832	0.956	2.157	1.215

[a] Presented as parameter estimate ± standard error from linear regression of held-out observations versus predictions.
[b] Root of the mean squared prediction errors.

expected (R^2 = 0.951, RMSPE = 2.010) as opposed to the rural areas (R^2 = 0.949, RMSPE = 2.010). Finally, looking at differences by season, we saw that the model predicts slightly better during the winter season than the summer across all years (R^2 = 0.944, RMSPE = 1.937 and R^2 = 0.879, RMSPE = 1.937 respectively).

Fig. 3 shows the spatial pattern of predicted *Ta* values from the *Ts* models, averaged over 2011 for the metropolitan New-York area. Mean predicted minimum *Ta* values for 2011 ranged from 10.87 to 13.35 °C. The figure shows how urban areas appear warmer than the surrounding areas. Also since we are predicting temperature at a high spatial resolution one can see the differences in yearly temperature between dense urban areas such as downtown NYC and nearby areas such as central park.

Table 2
Prediction accuracy: R^2 for model 2 temperature predictions (Final prediction model including locations without air/surface temperature for 2000–2011).

Yearly dataset	R^2
2000	0.957
2001	0.945
2002	0.950
2003	0.962
2004	0.926
2005	0.926
2006	0.935
2007	0.949
2008	0.948
2009	0.947
2010	0.936
2011	0.902
Mean 2000–2011	0.940

Table 3
Prediction accuracy: Ten-fold cross validated (CV) results for model 1 predictions for all states, sample cities, urban vs. rural and season analysis for a sample year (2005).

Type	CV R^2	Slope	RMSPE[a]
WV	0.945	0.960	1.738
VA	0.957	0.988	1.745
PA	0.952	0.987	1.798
MD	0.959	1.003	1.769
NY	0.945	1.002	2.159
RI	0.966	0.999	1.605
NJ	0.955	1.013	1.914
DE	0.957	0.994	1.426
CT	0.892	0.988	2.908
MA	0.943	0.998	1.687
VT	0.960	0.985	1.821
NH	0.944	1.001	2.238
ME	0.940	0.987	1.791
Boston	0.960	1.004	1.700
NYC	0.913	1.002	2.664
Rural	0.949	1.001	2.010
Urban	0.951	1.002	2.010
Winter	0.944	1.001	1.937
Summer	0.879	1.001	2.068

[a] Root of the mean squared prediction errors.

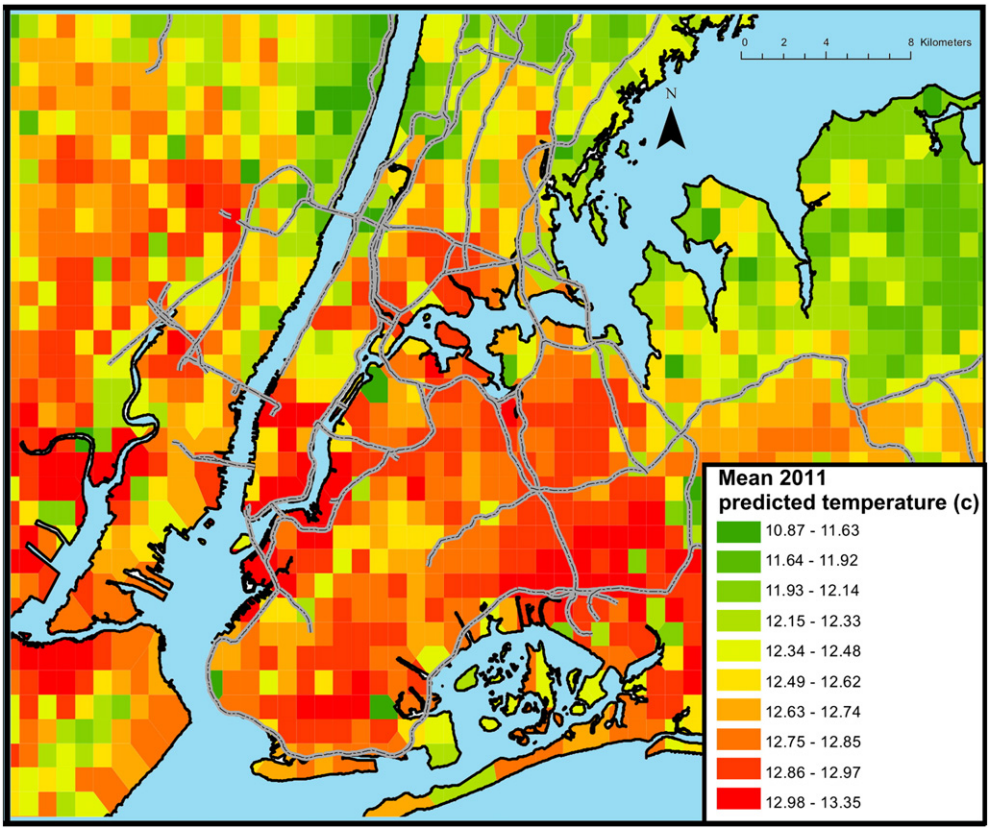

Fig. 3. The spatial pattern of predicted *Ta* values from the *Ts* models, averaged over the 2011 for the metropolitan New York area.

4. Discussion

In this paper we show that by using MODIS *Ts* data, one can reliably assess *Ta* across large regions. We demonstrate this across Northeastern USA during the years 2000–2011. This study improves on previous studies (Fu et al., 2011; Kloog et al., 2012; Vancutsem et al., 2010; Zhu et al., 2013) in a few key areas: First, daily calibrations of *Ts*–*Ta* results in a much better fit, since the association varies daily and hence results in better predictions. This daily calibration approach highlights the differences in short-term temperature values between grid cells. Second, our model allows us to predict daily *Ta* at a very high resolution, in large areas with varying geographical characteristics. Finally, as opposed to all previous studies, we calibrate the 24 h *Ta* with both night *Ts* and day *Ts* and present how both perform well with the daily calibration. We choose to use night *Ts* for our final calibration since the product is overall more stable (Vancutsem et al., 2010). Our results clearly present the strong association between *Ts* and *Ta*, describes very well both the intra and inter urban temporal and spatial patterns of *Ta*, and shows that *Ts* can be reliably used to predict *Ta* if modeled appropriately.

Comparison of our model to previous studies using the MODIS satellite to estimate *Ta* is not straightforward due to the wide span of spatial and temporal scales and the fact that, to the best of our knowledge, our model is the only model to predict daily temperature values even when *Ts* data is unavailable. Nonetheless, our model performance is excellent and in good agreement with all previous studies (Benali, Carvalho, Nunes, Carvalhais, & Santos, 2012; Fu et al., 2011; Kloog et al., 2012; Vancutsem et al., 2010; Zhu et al., 2013). For example, the prediction accuracy of our models (R^2 of 0.95–0.96, RMSPE = 2.157 °C) is significantly better than Zhu et al. (2013) (R = 0.86, RMSE = 3.43 °C) even when we predict for days *with no available Ts data* (R^2 of 0.94) and in good agreement with the Kloog study (Kloog et al., 2012) (R^2 = 0.94). These newly generated predictions from our model could be used in a variety of studies since they provide spatially–temporally resolved estimations of *Ta*, which are of great importance in areas with low station density or with highly variable spatial patterns between stations, e.g. mountainous regions, areas with large water bodies (such as lakes and rivers) and rural areas. Such applications could include hydrology, urban planning, agriculture, metrology, health related studies, among others. For example in epidemiology studies, such data could be used to assess the short and long term effect of exposure to temperature extremes. Such studies tend to use one central meteorological station to assign temperature exposure (Halonen et al., 2010). This introduces exposure error, and likely biases the effect estimates downward (Zeger et al., 2000). Our models will be critical for improving the accuracy of these health effect estimates in future epidemiological studies. Our predicted mean values are in a good agreement with the recorded annual mean minimum temperatures from *Ta* stations. For example, in NYC the annual mean difference between measured (NCDC, 2010) and predicted *Ta* was 0.35 °C (11.00°–11.35 °C respectively). In the Rural area of Harrisonburg, VA there was a slightly larger difference of 1.01 °C (11.98°–10.97 °C respectively).

There are three types of exposure measurement error in classic longitudinal studies of exposure and health (Zeger et al., 2000). The first error is derived from the difference between the daily personal exposures of each individual and the daily community average personal exposure. The second error stems from the difference between the daily community average personal exposure and the true ambient value. Finally, the third component of error is due the difference between the measured and the true ambient value (measurement error). We anticipate that by using our method which uses satellite MODIS *Ts* data to estimate individual level exposures, much of the second exposure error type will be minimized.

While the approach presented in this paper uses a hybrid model methodology incorporating remote sensing, advanced regressions and interpolations (and focuses on generating predictions for epidemiology) there are other advanced methodologies common in the climatology

field that are worth mentioning such as dynamical downscaling and sophisticate data assimilation (DA) techniques (Cocke & LaRow, 2000; Reichle, 2008; Wilby & Wigley, 1997). Dynamical downscaling takes output from a global or regional climate models (GCM/RCM) and converts the data into a regional, numerical models with much higher spatial resolution. There are multiple methods presented in recent years to reproduce surface temperature and precipitation fields including include dynamical modeling by nesting a regional climate model within a general circulation model (Cocke & LaRow, 2000) and climate-analog procedures (Houghton, 1996). DA methods combine data from measurements and models of earth into an optimal estimate of the geophysical fields of interest. DA systems interpolate and extrapolate the remote sensing observations and provide complete estimates at the scales required by the application, both in time and in the spatial dimensions (Reichle, 2008). Among the recent applications of DA are geophysical processes, detection of changes and trends in the earth system and forecasting. DA uses sophisticated algorithms that can vary greatly since conditions in different regions and between ocean and land vary considerably. Our presented models while not as advanced are much simpler to construct and run (to generate daily predictions) and thus more suitable for epidemiology studies.

There are some limitations in the present study. For one, since we use a daily calibration method, it requires continuous, relatively high spatial resolution daily *Ta* stations which are not always available in other regions or areas of the world. Also, the spatial distribution of *Ta* stations may change dramatically from year to year in some areas. In the presented study, there were as few as 513 operating *Ta* stations in 2000 and as many as 2860 during 2011. Though this does effect the overall performance of the model ($R^2 = 0.933$ and RMSPE of 2.109 for 2000 vs $R^2 = 0.955$ and RMSPE of 2.083 for 2011), it suggests that the model can predict reliably even with lower spatial distribution of stations. A final limitation is that *Ts* data from MODIS only starts on March 2000, limiting how far back one can predict *Ta*.

In conclusion, we demonstrate how *Ts* can be used reliably to predict daily *Ta* at high resolution in large geographical areas even in days with missing satellite data. Our results could be used in a variety of studies since they provide spatially–temporally resolved estimations of *Ta*.

Acknowledgments

This study was funded by the Harvard EPA PM Clean Air Research Center (CLARC) (R-834798), NIEHS ES000002, and the following R21 climate grants ES020695 and AG040027. The authors also thank the www.weatherunderground.com service and in particular Brendan Hayes and John Celenza for their help.

Appendix

Table A1
Prediction accuracy: Ten-fold cross validated R^2 for air temperature stage 1 predictions (calibration stage using MODIS day surface temperature for 2000–2011).

Yearly dataset	CV R^2	Intercept[a]	Slope[a]	CV R^2 Spatial	CV R^2 Temporal	RMSPE[b]	RMSPE[b] Spatial
2000	0.937	0.002 ± 0.02	1.001 ± 0.001	0.828	0.951	2.133	1.224
2001	0.947	0.017 ± 0.017	1.001 ± 0.001	0.765	0.957	2.086	1.294
2002	0.955	0.014 ± 0.017	0.999 ± 0.001	0.839	0.965	2.121	1.175
2003	0.961	0.02 ± 0.016	0.999 ± 0.001	0.893	0.967	2.088	1.155
2004	0.953	0.027 ± 0.014	0.999 ± 0.001	0.864	0.961	2.157	1.223
2005	0.951	0.027 ± 0.013	0.999 ± 0.001	0.843	0.960	2.305	1.169
2006	0.949	−0.004 ± 0.012	1.00 ± 0.001	0.807	0.958	2.051	1.143
2007	0.946	−0.025 ± 0.012	1.001 ± 0.001	0.952	0.954	2.339	1.230
2008	0.953	0.003 ± 0.011	0.999 ± 0.001	0.831	0.957	2.056	1.272
2009	0.952	0.003 ± 0.009	1.000 ± 0.001	0.799	0.958	2.188	1.298
2010	0.952	−0.001 ± 0.009	1.000 ± 0.001	0.847	0.959	2.131	1.240
2011	0.960	−0.006 ± 0.006	1.000 ± 0.000	0.873	0.967	2.023	1.084
Mean 2000–2008	0.951	0.006 ± 0.013	1.000 ± 0.001	0.845	0.959	2.140	1.209

[a] Presented as parameter estimate ± SE from linear regression of held-out observations versus predictions.
[b] Root of the mean squared prediction errors.

References

Basu, R., Feng, W. Y., & Ostro, B.D. (2008). Characterizing temperature and mortality in nine California counties. *Epidemiology*, *19*, 138.
Benali, A., Carvalho, A.C., Nunes, J. P., Carvalhais, N., & Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sensing of Environment*, *124*, 108–121, http://dx.doi.org/10.1016/j.rse.2012.04.024.
Cocke, S., & LaRow, T. E. (2000). Seasonal predictions using a regional spectral model embedded within a coupled ocean–atmosphere model. *Monthly Weather Review*, *128*.
Dousset, B. (1989). *AVHRR-derived cloudiness and surface temperature patterns over the Los Angeles area and their relationship to land use, 2132–2137.*
Florio, E. N., Lele, S. R., Chang, Y. C., Sterner, R., & Glass, G. E. (2004). Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: A statistical approach. *International Journal of Remote Sensing*, *25*, 2979–2994.
Fu, G., Shen, Z., Zhang, X., Shi, P., Zhang, Y., & Wu, J. (2011). Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. *Acta Ecologica Sinica*, *31*, 8–13.
Garske, T., Ferguson, N. M., & Ghani, A.C. (2013). Estimating air temperature and its influence on malaria transmission across Africa. *PLoS ONE*, *8*, e56487, http://dx.doi.org/10.1371/journal.pone.0056487.
Goetz, S. J., Prince, S. D., & Small, J. (2000). Advances in satellite remote sensing of environmental variables for epidemiological applications. *Advances in Parasitology (S.E.R. S.I. Hayed.), Vol. 47.* (pp. 289–307). : Academic Press.
Halonen, J. I., Zanobetti, A., Sparrow, D., Vokonas, P.S., & Schwartz, J. (2010). Associations between outdoor temperature and markers of inflammation: a cohort study. *Environmental Health*, *9*, 42.
Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national landcover database for the United States. *Photogrammetric Engineering and Remote Sensing*, *70*, 829–840.
Houghton, J. T. (1996). *Climate change 1995: The science of climate change: Contribution of working group I to the second assessment report of the Intergovernmental Panel on Climate Change.* : Cambridge University Press.
Jensen, J. R. (2009). *Remote sensing of the environment.* Pearson Education India.
Kloog, I., Chudnovsky, A., Koutrakis, P., & Schwartz, J. (2012). Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA. *Science of the Total Environment*, *432*, 85–92, http://dx.doi.org/10.1016/j.scitotenv.2012.05.095.
Lennon, J. J., & Turner, J. R. G. (1995). Predicting the spatial distribution of climate: Temperature in Great-Britain. *Journal of Animal Ecology*, *64*, 370–392.
Maune, D. (2007). Digital elevation model technologies and applications: The DEM users manual. *American Society for Photogrammetry and Remote Sensing*, p. 620, ISBN 1-57083-082-7 (Maune, DF).
Medina-Ramón, M., Zanobetti, A., Cavanagh, D. P., & Schwartz, J. (2006). Extreme temperatures and mortality: Assessing effect modification by personal characteristics and specific cause of death in a multi-city case-only analysis. *Environmental Health Perspectives*, *114*, 1331–1336, http://dx.doi.org/10.1289/ehp.9074.
NCDC (2010). *The National Climatic Data Center data inventories.*
Ostro, B.D., Roth, L. A., Green, R. S., & Basu, R. (2009). Estimating the mortality effect of the July 2006 California heat wave. *Environmental Research*, *109*, 614–619, http://dx.doi.org/10.1016/j.envres.2009.03.010.
Prihodko, L., & Goward, S. N. Z. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*, *60*, 335–346.
Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, *31*, 1411–1418, http://dx.doi.org/10.1016/j.advwatres.2008.01.001.
Ren, C., O'Neill, M. S., Park, S. K., Sparrow, D., Vokonas, P., & Schwartz, J. (2011). Ambient temperature, air pollution, and heart rate variability in an aging population. *American Journal of Epidemiology*, *173*, 1013–1021, http://dx.doi.org/10.1093/aje/kwq477.
Smith, W. L., Leslie, L. M., Diak, G. R., Goodman, B.M., Velden, C. S., Callan, G. M., et al. (1988). The integration of meteorological satellite imagery and numerical dynamical forecast

models. *Philosophical Transactions of the Royal Society A — Mathematical Physical and Engineering Sciences*, *324*, 317–323, http://dx.doi.org/10.1098/rsta.1988.0022.

Vancutsem, C., Ceccato, P., Dinku, T., & Connor, S. J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, *114*, 449–465, http://dx.doi.org/10.1016/j.rse.2009.10.002.

Vicente-Serrano, S., Saz-Sánchez, M., & Cuadrat, J. (2003). Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature. *Climate Research*, *24*, 161–180, http://dx.doi.org/10.3354/cr024161.

Von Klot, S., Melly, S., Coull, B., Dutton, J., & Schwartz, J. (2008). Daily temperature at residence vs central measurements in eastern Massachusetts. *Epidemiology*, *19*, S210.

Wan, Z. (2008). New refinements and validation of the MODIS land-surface temperature/emissivity products. *Remote Sensing of Environment*, *112*, 59–74.

Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, *34*, 892–905.

Wang, L., Koike, T., Yang, K., & Yeh, P. J. -F. (2009). Assessment of a distributed biosphere hydrological model against streamflow and MODIS land surface temperature in the upper Tone River Basin. *Journal of Hydrology*, *377*, 21–34, http://dx.doi.org/10.1016/j.jhydrol.2009.08.005.

Watson, R. T., & Albritton, D. L. (2001). *Climate change 2001: Synthesis report: Third assessment report of the Intergovernmental Panel on Climate Change.* : Cambridge University Press.

Weng, Q., Lu, D., & Schubring, J. (2004). Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment*, *89*, 467–483.

Wilby, R. L., & Wigley, T. M. L. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, *21*, 530–548.

Yang, J. S., Wang, Y. Q., & August, P. V. (2004). Estimation of land surface temperature using spatial interpolation and satellite-derived surface emissivity. *Journal of Environmental Information*, *4*, 40–47.

Zanobetti, A., & Schwartz, J. (2008). Temperature and mortality in nine US cities. *Epidemiology*, *19*, 563.

Zeger, S. L., Thomas, D., Dominici, F., Samet, J. M., Schwartz, J., Dockery, D., et al. (2000). Exposure measurement error in time-series studies of air pollution: Concepts and consequences. *Env. Health Perspectives*, *108*, 419–426.

Zhang, L., Huang, J., Guo, R., Li, X., Sun, W., & Wang, X. (2013). Spatio-temporal reconstruction of air temperature maps and their application to estimate rice growing season heat accumulation using multi-temporal MODIS data. *Journal of Zhejiang University. Science. B*, *14*, 144–161, http://dx.doi.org/10.1631/jzus.B1200169.

Zhu, W., Lű, A., & Jia, S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, *130*, 62–73, http://dx.doi.org/10.1016/j.rse.2012.10.034.